社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
EA2002-96 (2002-12)

短時間自己相関処理を用いた騒音環境下の音声区間検出

兒玉 里美*†, 寺尾 和也†, 北尾 英樹††, 岩田 収††, 中村 正孝†

† 広島工業大学
〒731-5193 広島県広島市佐伯区三宅 2-1-1
†† 富士通テン株式会社技術開発部
〒652-8510 神戸市兵庫区御所通 1-2-28

E-mail: † nakamura@cc.it-hiroshima.ac.jp, †† kitao@ten.fujitsu.co.jp

あらまし 本文では、実騒音環境下で雑音耐性の強い音声区間検出法の実現を目標として、短時間自己相関処理を用いて雑音中の音声に関連した信号を取り出し、その二乗振幅レベルの閾値判別による音声区間検出法を提案する。ここで、より正確な音声区間検出のために、非音声区間でのノイズレベル評価とそれを用いた閾値レベルの新しい自動設定法を開発している。また、音声区間の継続時間を正誤判断に用いて、非定常雑音の耐性を高め、さらに、音声発声のはじまりと終りの部分での低レベル並びに、促音などによる単語の中での著しく低いレベルに対しては、区間検出パルスの前側と後側にパルス幅を付加する処理によって、それらの区間検出を可能としている。
キーワード 音声信号処理、音声区間、自己相関、雑音耐性、閾値処理、閾値レベルの自動設定

A Study on Speech Period Detection by Short-Time Auto-Correlation Processing under Noisy Environments

Satomi Kodama†, Kazuya Terao†
Hideki Kitao††, Osamu Iwata††, and Masataka Nakamura†

† Hiroshima Institute of Technology
2-1-1 Miyake, Saeki-ku, Hiroshima 731-5193 Japan
†† Fujitsu Ten Limited
1-2-28 Gosyodori, Hyogo-ku, Kobe 652-8510 Japan

Abstract The purpose of this study is to develop the algorism for robust speech period detection under real noisy environments. In this paper, we propose a speech period detection method in which the related signals to speech buried in noise is extracted by using the short-time autocorrelation processing and its square values are discriminated by a certain threshold level. In order to achieve more accurate speech period detection, a new method for estimating the noise level during an unvoiced period and for setting automatically a threshold level based on the estimated noise is developed in this study. And also, the true-false decision of detecting periods is provided for the purpose of eliminating the output due to a non-stationary noise. Furthermore, in consideration of the error detection caused by extremely low level of speech around at beginning and ending of speech as well as at midway through the pronunciation of a word, the output pulse width of detecting periods is designed to be increased at its front and rear.

Key Words speech signal processing, speech period, autocorrelation, robustness against noise, thresholding, automatic threshold level setting

1. まえがき

音声入力マンマシンインターフェイスとしての音声認識システムの入力部において、音声区間を検出することが重要である。区間検出法に音声エネルギーの閾値と零交差回数を判断に用いる方法がある。しかし、その方法では、人ごみや走行時の自動車室内など実環境下で、音声信号に非定常雑音が重畳されていることが多く、低 S/N の場合、著しく検出率が劣化する。そのため、騒音を考慮しての区間検出システム並びに受音段階で騒音成分を抑えた区間検出など多くの研究がなされてきた。しかしながら、現在なお、実騒音環境下での正確な音声区間検出は困難な課題のひとつである。

本研究では、音声エネルギーそのものでなく騒音中から音声相当分を効果的に抽出できる短時間自己相関とその自乗値の閾値判断による区間検出を検討している。ところで、閾値判断には、話者並びに周囲の環境などにより、音声レベルと雑音レベルがその時々に著しく異なるため、閾値を一定値に設定することができないという問題がある。

そこで、本文では短時間自己相関自乗値に対する閾値レベルを、非音声区間の自乗値を計測し、それを用いて閾値の自動設定する方法を提案する。次に、語頭・語尾、並びに単語中の促音などの信号レベルが極端に低下した部分では区間検出が不可能となるという問題がある。本文では、その対策として、区間検出パルスの前方だけでなく、前方後方のパルス幅を伸長する処理法を示す。おわりに、本音声区間検出の全処理システムを示す。

2. 短時間自己相関

音声データ x(n)に対する短時間自己相関関数 C(n)を次式に示す。N は相関個数、M は隔たり時間である。

$$C(n)=\frac{1}{N}\sum_{i=0}^{N-1}x(n-i)\cdot x(n-i-M) \qquad (1)$$

本システムでは準実時間処理を目標とするため、あらためて短時間自己相関 C(n)を次の式(2)のように定義する。

$$C(n)=\frac{1}{N}\sum_{i=-N+1}^{0}x(n-i)\cdot x(n-i-M) \qquad (2)$$

ここでは、相関個数 N を 5 個、隔たり数 M を 66(6msec)としている。図 1 に式(2)に示す短時間自己相関処理を図示する。図 1 に示すように C(n)は 1 サンプル毎に得る。図 2 は、自動車走行中に女性が"動物園"と発声しているデータ例である。録音時の環境は中型車においてエアコンを Low にして走行状態の車中である。なお、データはサンプリング周波数 11.025kHz、量子数 16bit、モノラルである。

短時間自己相関処理前に、音声データ中のエアコンの雑音を除去するため IIR 型 3 次 High Pass Filter を用い 300Hz 以下の成分を除去する。図 2 のデータに前述のフィルタ処理をした後の短時間自己相関結果を図 3 に示す。

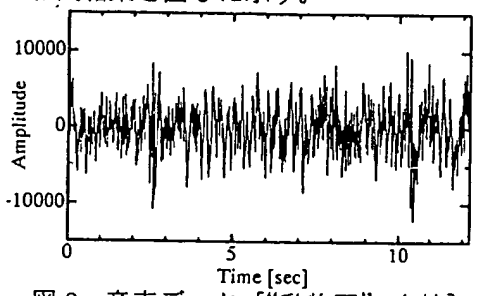


図 2　音声データ ["動物園" /女性]

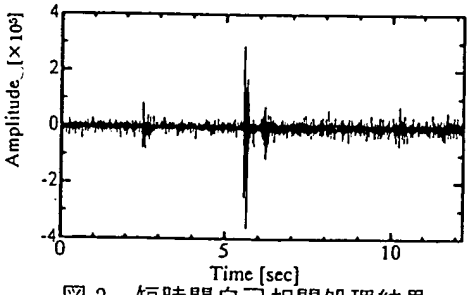


図 3　短時間自己相関処理結果

3. 区間検出と閾値レベル

本研究では、次式の短時間自己相関の 2 乗平均 p(n)を対象に区間検出を考える。

$$p(n)=\sqrt{\frac{1}{128}\sum_{j=-127+n}^{n}C^2(j)} \qquad (3)$$

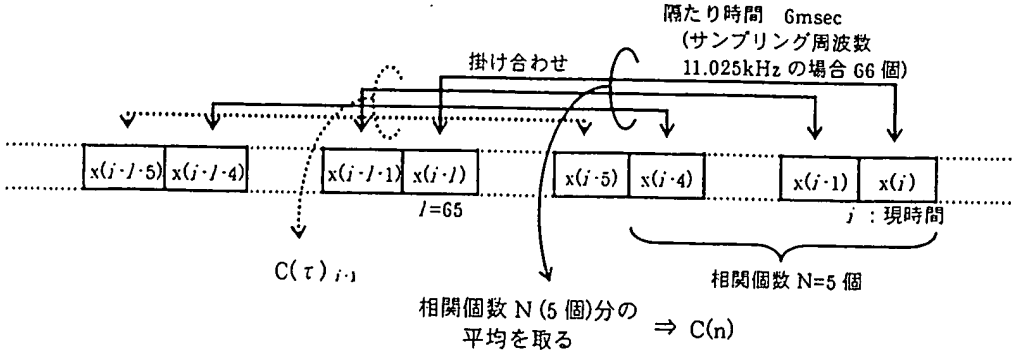


図 1　本システムにおける短時間自己相関処理

短時間自己相関自乗平均値p(n)に対して図4に示すように、ある閾値に対して音声区間・非音声区間の判定を行う。本システムでは、区間検出の閾値は、非音声区間の短時間自己相関の二乗平均p(n)を測定・評価して、それにある定数を掛けた値に自動設定する。

すなわち、閾値は、以下のような処理手順によって自動的に設定する。

[1]音声区間パルスが零の場合(図4：a区間)、

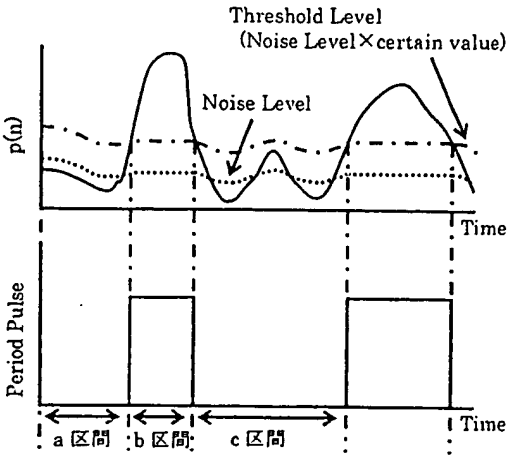


図4　閾値と音声区間パルス

式(4)に示すように現時間から過去250msec分のp(n)の平均値を求める（以下 Noise Level をNL(n)と記す)。

各1サンプル毎にNL(n)を求める。

$$NL(n) = \frac{1}{2756} \sum_{k=-2755}^{0} p(n-k) \qquad (4)$$

そして、算出されたNL(n)にあらかじめ定めた係数（ここでは1.5）を掛け、次のサンプル点での閾値とする。

[2]音声区間検出パルス(以下出力ゲートという。)が出力された場合(図4：b区間)、NL(n)は、出力ゲートが出力される直前の値を保持持続する。

[3]再び検出パルスが立ち下がると(図4:c区間)、処理[1]を繰り返す。

以上のように図2の音声データのNL(n)に対する閾値自動設定処理結果を図5に、また、その場合の出力ゲートを図6に示す。なお、NL(n)の測定は非音声区間に行うことから、図5に示すように初期の閾値は高い値に設定する必要がある。

4. 検出パルスの正誤判断とパルス前後付加

4.1 継続時間判別処理

図5を観察すると、騒音のp(n)は急峻なパルス

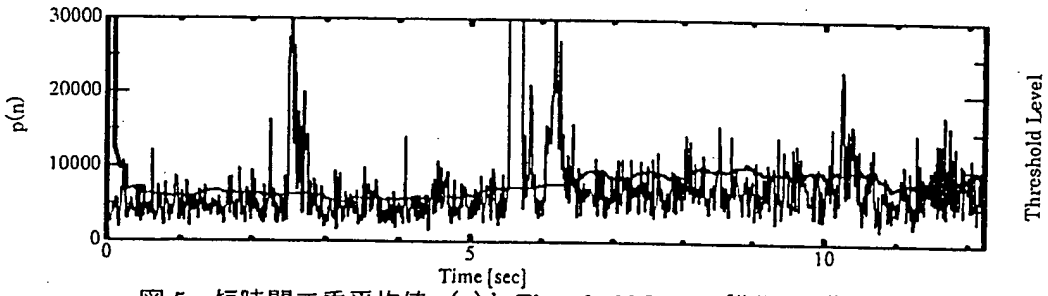


図5　短時間二乗平均値p(n)とThreshold Level［“動物園”／女性］

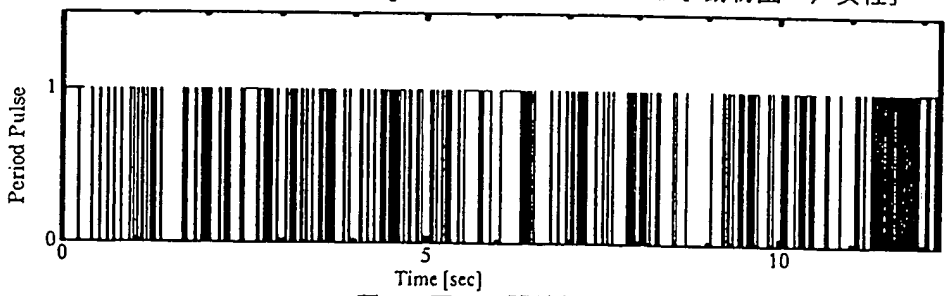


図6　図5の閾値処理結果

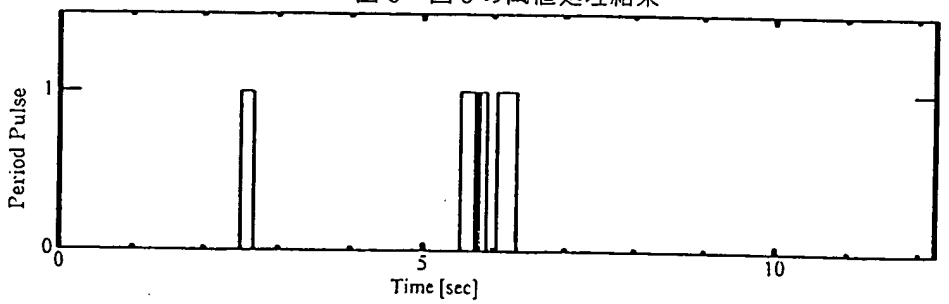
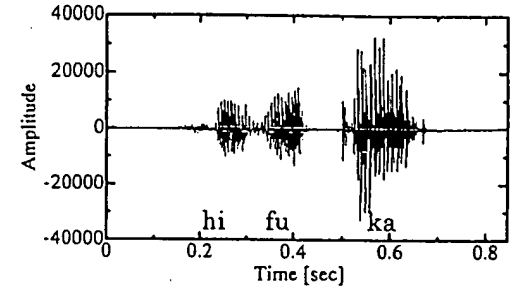


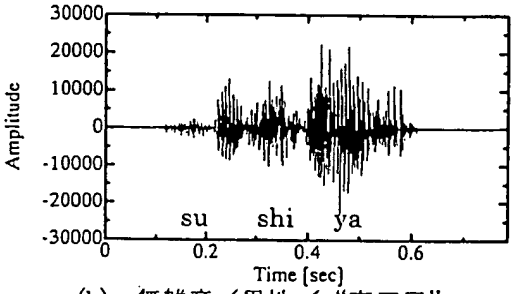
図7　継続時間判別処理結果

状変化を呈し、閾値を超え、誤った検出パルス出力が多数現れている。この急峻なパルス状変化は、平滑処理(移動平均)によっても除去できるが、同時に信号に相当するp(n)も平滑され、区間検出があいまいとする。

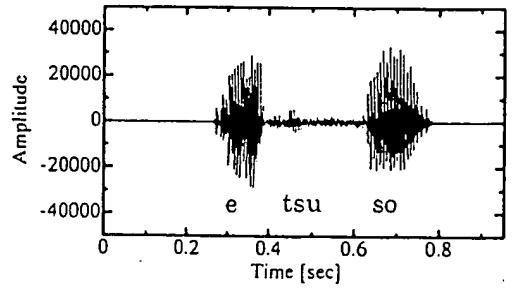
そこで、本研究では突発的なパルス状変化による出力ゲートは、継続時間が短いことを判断に用いて区間検出パルスとしては出力しない処置を講ずる。なお、ここではこの判別の継続時間(サンプル数kに換算)を60msecとしている。以上のような処理を追加した場合のゲート出力結果を図7に示す。図6と比較すると、図7の結果では、パルス状変化に対する出力ゲートは生じなくなる。この継続時間判別処理はとくに自動車室内におけるワイパー音、ウインカー音による誤検出が、よく除去できており、継続時間の短い非定常騒音の棄却に有効である。



(a) 無雑音／男性／“皮膚科”



(b) 無雑音／男性／“寿司屋”



(c) 無雑音／男性／“エッソ”

図8 鼻音、摩擦音、促音を含む音声波形例(無雑音)

4.2 ゲート前後付加処理

ここまでの処理によって得られた出力ゲート間の音声データを実際に聴取すると、途中で途切れたり、語頭・語尾等が正確に聞き取れない結果となった。これは、発声開始時と終了直前に音声レベルが著しく低くなることがひとつの原因である。さらに、図8の例に示すように、音韻を特徴づける先行部あるいは、単語中の摩擦音(ひ、しぃー)、促音(っ)のレベルが極めて低いところに原因している。

実騒音環境下では、それらは−15dB以下のS/Nになる。この音韻認識に重要な、極めて微弱な信号に対する区間検出のため、本システムでは、区間検出パルスに前後にパルス幅を付加する処理をもたせている(ここでは120msec.とする)。これは、発声開始時・終了直前だけでなく、単語中での極めて低いレベルの部分の区間検出の補間に役立つ。

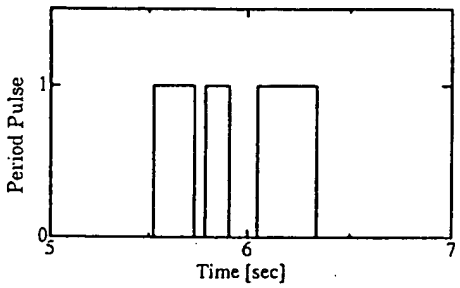


図9 継続時間判別後の出力ゲート

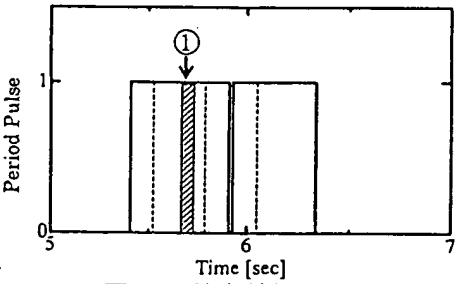


図10 前方付加処理結果

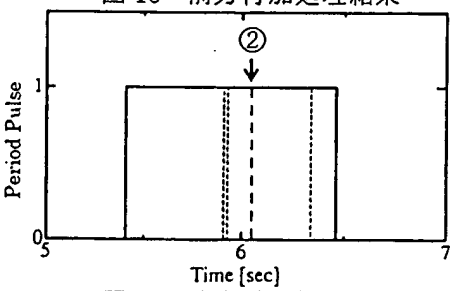


図11 後方付加処理結果

区間検出パルズの前後にパルス幅を付加する様子を図9〜図11に示す。図9に図7の継続時間判別結果後のゲート出力の横軸を拡大して再掲する。図10、図11に図9のゲート出力に前方・後方付加処理した結果を示す。図10において点線は図9のゲート出力を表す。なお、本システムでは継続時間判別処理とこの前方付加処理を同時に行なっている。すなわち、閾値処理結果において1サンプル前のデータと現データを比較し、

現データの方が大きければカウント処理を始める。カウントが継続判別時間と等しくなったとき、判別時間と前方付加時間を加えた時間だけ遡及し、現時間までゲートを出力する。図 10 中の①部分は、前処理で現れた出力ゲートに付加パルスが重畳している。

図 11 にゲート出力の後方付加処理結果を示す。点線は図 10 の結果を表す。前方向付加処理結果において、1 サンプル前のデータと現時点のデータを比較し、現データのほうが小さくなると図 11 中の②の時点まで後方付加を行なう。この場合は遡及処理を行なう必要はない。それぞれ前方付加処理、後方付加処理の流れを付録に示す。

5. 全処理システムの流れと音声区間検出結果

本音声区間検出システムの全体の処理の流れを図 12 に示す。

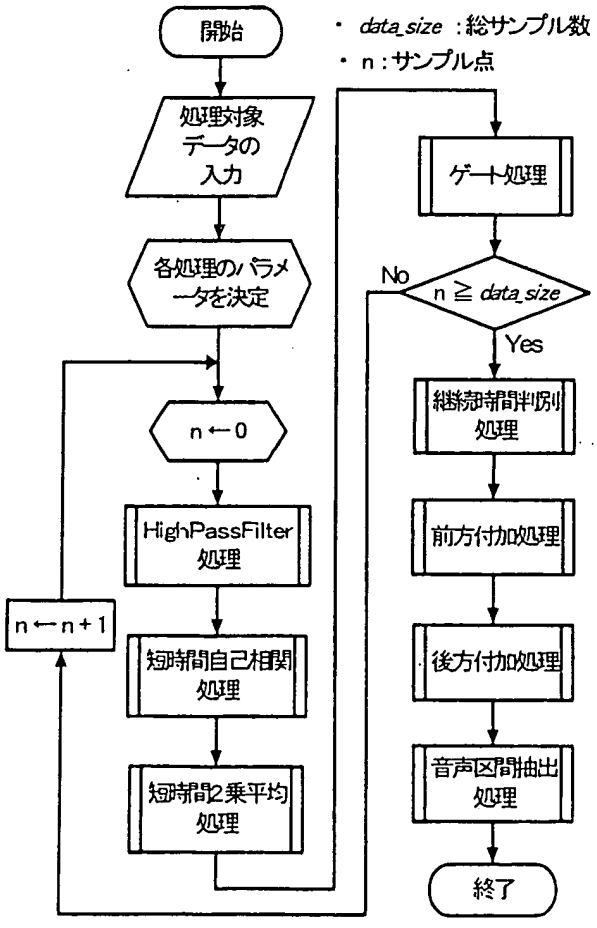


図 12　システム全体の流れ

中型車（エアコンは Mid）の走行中の車内での騒音を含む女性の発声例に対する本システムの区間検出パルス結果を図 13 に掲げる。各図の上段は雑音を含む原信号、下段は 300Hz の High Pass Filter 通過出力波形である。また、上段での細い実線が、パルス幅を付加しない場合、太い

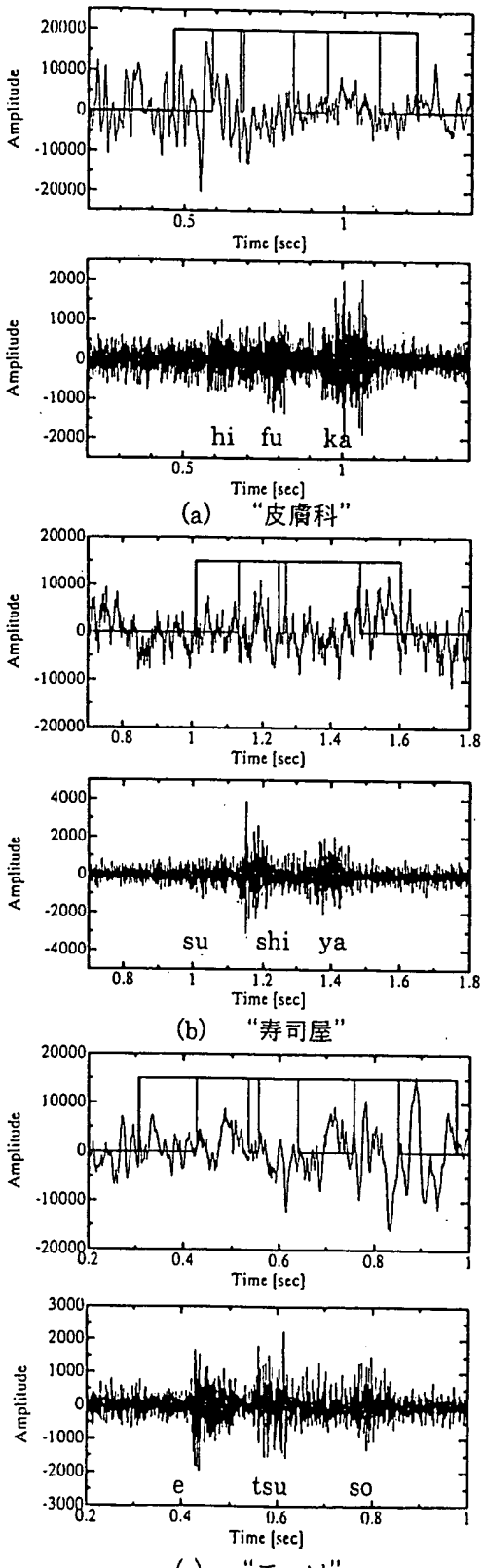


図 13　区間パルス検出結果
上段：原信号と区間検出
下段：HPF 処理後

実線は、前後にパルス幅を付加した出力ゲートを表す。この結果、実線の出力ゲートで示すように正確な聴取テスト結果が得られ、語頭・語尾、また単語中の非常にレベルの低い部分も良好な区間検出可能となっている。

6. あとがき

本文では短時間自己相関を用いた雑音にロバストな音声区間検出の一方式について述べた。その中で、短時間自己相関の自乗値に対する閾値判別に関して、新しい閾値レベルの自動設定法を提案した。

騒音による短時間自己相関自乗平均値は全く零とはならない。それによる急峻な変化は、幅の狭い誤区間検出パルスを発生する。これに対して平滑処理を用いると、反対に信号出力ゲートに誤差を生ずる。それ故、本研究では、出力ゲートの継続時間判断を用いて幅の狭い誤検出を除去している。これは、ワイパー音、ウインカー音などの非定常性騒音の望まない検出を除去可能としている。

さらに、実騒音環境下で、極めて低 S/N となる語頭・語尾、並びに単語中の促音などの部分の音声区間検出が困難なことを指摘した。その対策として、本文では区間検出パルスの前方だけでなく後方にもパルス幅を付加する処理法を示した。その結果、本システムによって良好な区間検出結果が得られた。

しかし、S/N が極めて悪い場合(音声レベルの低い女声に多い)、区間検出は十分に満足する結果とはなっていない。従って、ピンクノイズの背景騒音をさらによく抑制した短時間自己相関結果が得られる手法の開発が今後の大きな課題である。

参考文献

[1] 大賀寿朗, 山崎芳男, 金田豊, "音響システムとディジタル処理", 電子情報通信学会, 1995.

[2] 鹿野清宏, 中村哲, 伊勢史郎, "音声・音情報のディジタル信号処理", 昭光堂, 1997.

[3] 金田豊, "マイクロホンアレーを用いた雑音下での音声区間検出", 電子情報通信学会論文誌, VoL.J80-A no.8, pp.1391-1398, 1990.

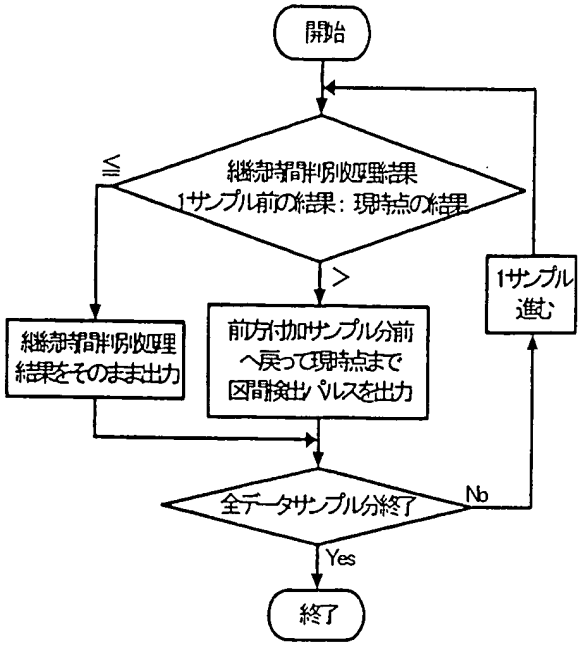
[4] 内藤正樹, 黒岩眞吾, 山本誠一, 武田一哉, "部分文仮説のゆう度を用いた連続音声認識のための音声区間検出法", 電子情報通信学会論文誌, VoL.J80-D-II no.11, pp.2895-2903, 1997.

[5] 渡部生聖, 山田武志, 浅野太, 北脇信彦, "環境音モデルと HMM 合成を用いた音声区間検出の検討", 信学技報, SP2000-84, pp.55-60, 2000.

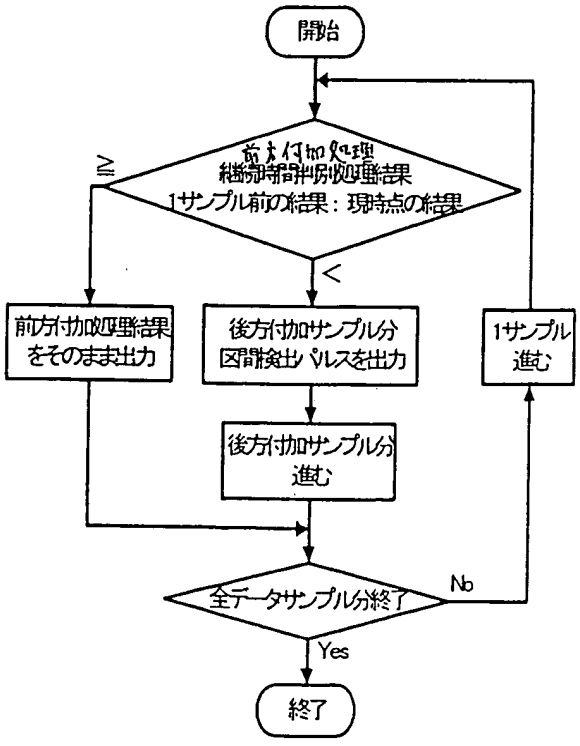
[6] 藤本雅清, 有木康雄, "マイクロフォンアレイとカルマンフィルタを用いたノイズロバストなハンズフリー音声認識の検討", 信学技報, SP2002-9, pp.13-18, 2002.

[7] 中村哲, "実音響環境に頑健な音声認識を目指して", 信学技報, SP2002-12, pp.31-36, 2002.

付録



A-1　前方付加処理の流れ



A-2　後方付加処理の流れ